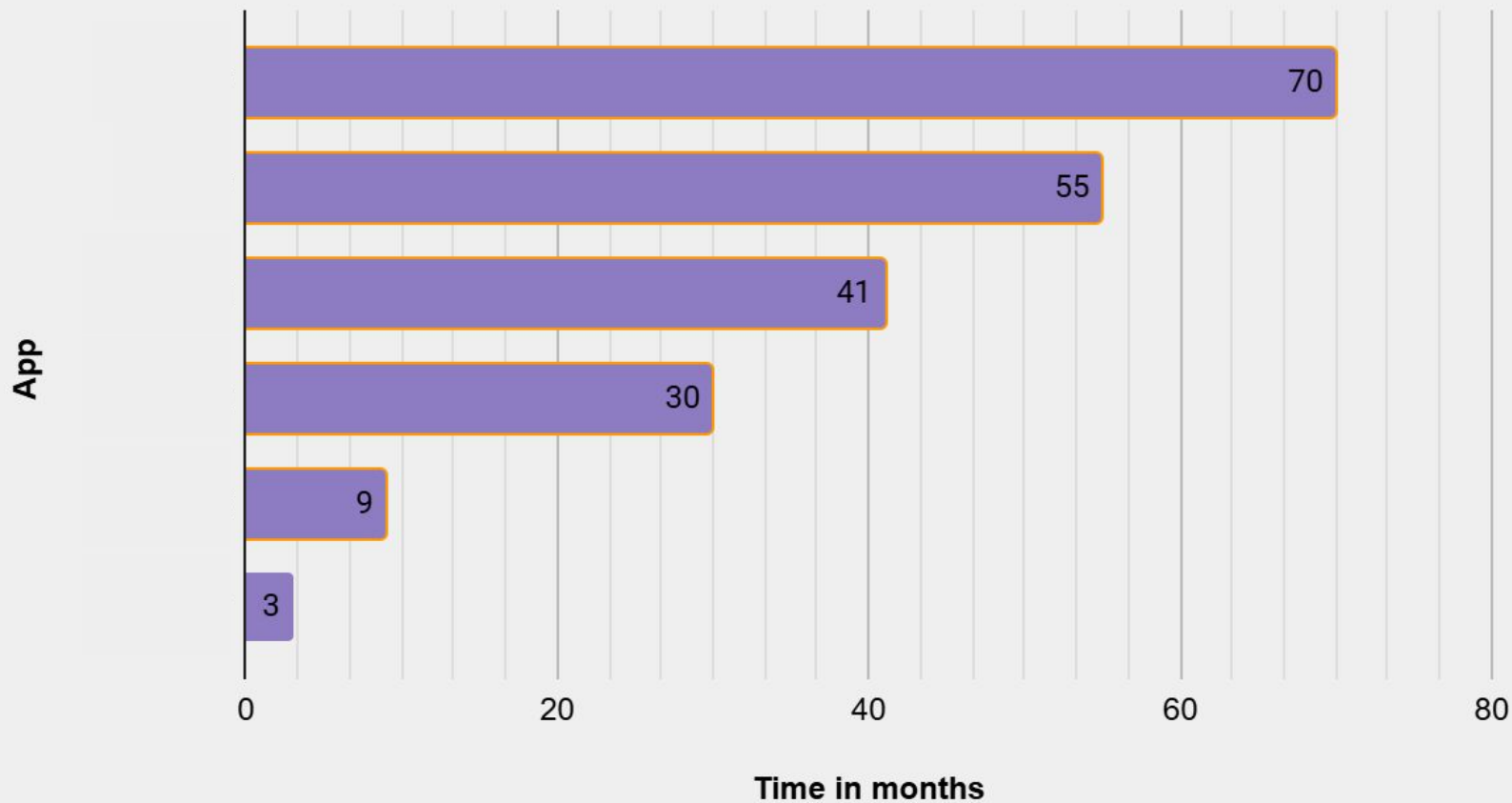


The Prompting Playbook: Strategies for Winning with Large Language Models

Time for Apps to reach 100M Monthly Users



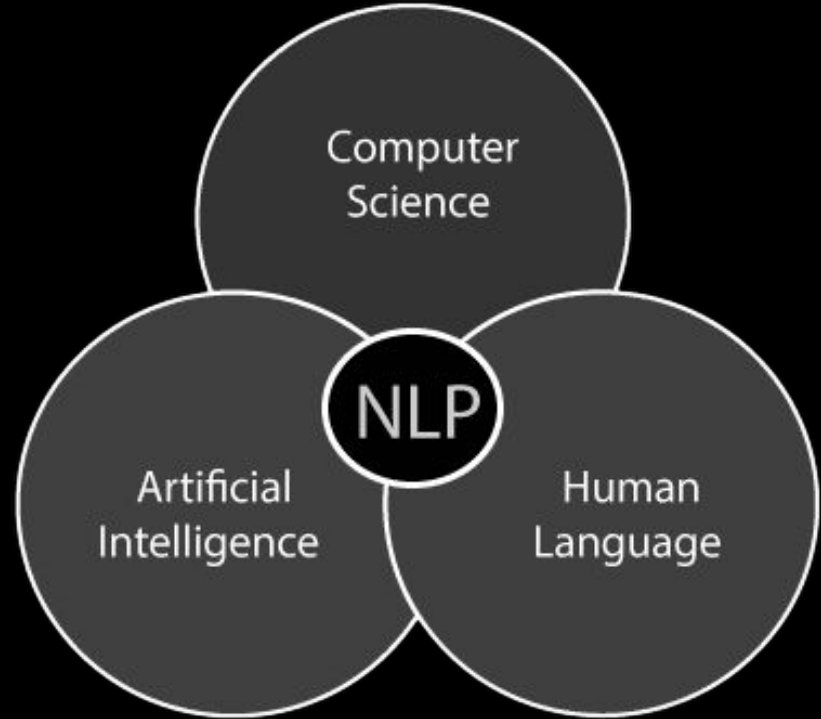
Generative Pretrained Transformer

- GPT is a transformer-based model that is trained on a large dataset of text to generate human-like text.
- OpenAI's ChatGPT is refined using Reinforcement Learning from Human Feedback (RLHF)




What is prompt engineering?

Concept in Natural Language Processing (NLP) that involves embedding

descriptions of tasks in input to prompt the model to output the desired results.



Everyday usage examples

AI Type			
Chatbot Customer Service	"Help me"	How do I reset my password"	"Instructions for password reset"
Medical Diagnosis AI	"I don't feel well"	"I have a fever and sore throat"	"Symptoms: fever, sore throat"
Content Creation	"Write something about climate change"	"Write a 500-word essay on climate change impacts"	"500-word essay, climate change impacts"
Financial Analysis	"Tell me about stocks"	"What's the current value of Apple stocks?"	"Current value, Apple Stocks?"

Elements of a prompt

01

Be specific

02

Use Concise
Language

03

Avoid
leading
questions

04

Apply
structure

05

Give
examples

06

Do iterate
and refine

**What elements
make up a
successful
prompt?**

07

Provide response
length

Types of prompts

Zero-shot	Used to interact with the model and will not contain examples
Few-shot	Enables in-context learning where human provides demonstrations in prompt to steer the model to better performance
Chain-of-thought	Enables complex reasoning capabilities through intermediate reasoning steps.

Zero-shot: Personal Development & Community Involvement

- Time Management & Study Skills:

"Describe a time management strategy that can help community college students balance work, school, and personal life."

- Civic Responsibility & Social Change:

"What is one local policy or social issue in Milwaukee that you think should be changed, and what steps would you take to advocate for it?"

Few-shot: Summarization with a focus

Prompt:

Summarize the following meeting notes focusing only on action items assigned to 'Sarah'.

Notes: "Meeting recap: Discussed Q3 goals. John to draft proposal. Sarah needs to update the client presentation by Friday. Reviewed budget - looks okay. Team lunch next week."

Summary: Sarah to update client presentation by Friday.

Notes: "Project Alpha sync: Code freeze next week. Mark to test deployment script. Sarah will finalize the user documentation. Lisa is handling the demo."

Summary: Sarah will finalize the user documentation.

Notes: "Marketing brainstorm: Ideas for new campaign discussed. Peter assigned to research competitors. Sarah volunteered to create the initial ad copy drafts by Wednesday. Need to finalize slogan."



Summary:

Few-shot: Specific style text generation

Prompt:

Write a short, enthusiastic product announcement using emojis.

Product: New Running Shoes

Announcement: Step up your game!  Our new FeatherLite runners are here! ✨ So comfy & fast!  Get yours today! 🎉

Product: Coffee Mug

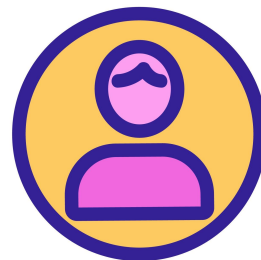
Announcement: Wake up happy!  Our new ceramic mugs are perfect for your morning brew! ☕ Grab one now! 😊

Product: Wireless Headphones

Announcement:

CoT: Task example

- **Prompt:** Q: I need to paint a rectangular room. The room is 12 feet long, 10 feet wide, and 8 feet high. I don't need to paint the ceiling or the floor. One gallon of paint covers 400 square feet. How many gallons of paint do I need? *Let's think step-by-step.*
- *(Task: Breaking down a geometry/rate problem into sub-problems: calculate wall area, then divide by coverage)*



What is your name?

I'm Gemini, a large language model, or an artificial intelligence, if you prefer.

Right now, you're trying out a new experimental model as an early preview of my latest developments. It brings significant improvements, even if it doesn't always work as expected while in experimental state.

What can I do for you?



Interacting with LLMs

System Message

User Message

Assistant response

###You are
a friendly
assistant
called,
Marco.###



What is
your
name?



My name is
Marco, nice
to meet you

System Message

###You are a friendly assistant called Marco. You always respond in a rhyme.###



What is your name?



If you need a hand, just say the word, you see, My name is Marco, friendly as can be!

System Message

###You are a friendly assistant called Marco. You always respond in a rhyme.###



What is your name?



If you need a hand, just say the word, you see,
My name is Marco, friendly as can be!



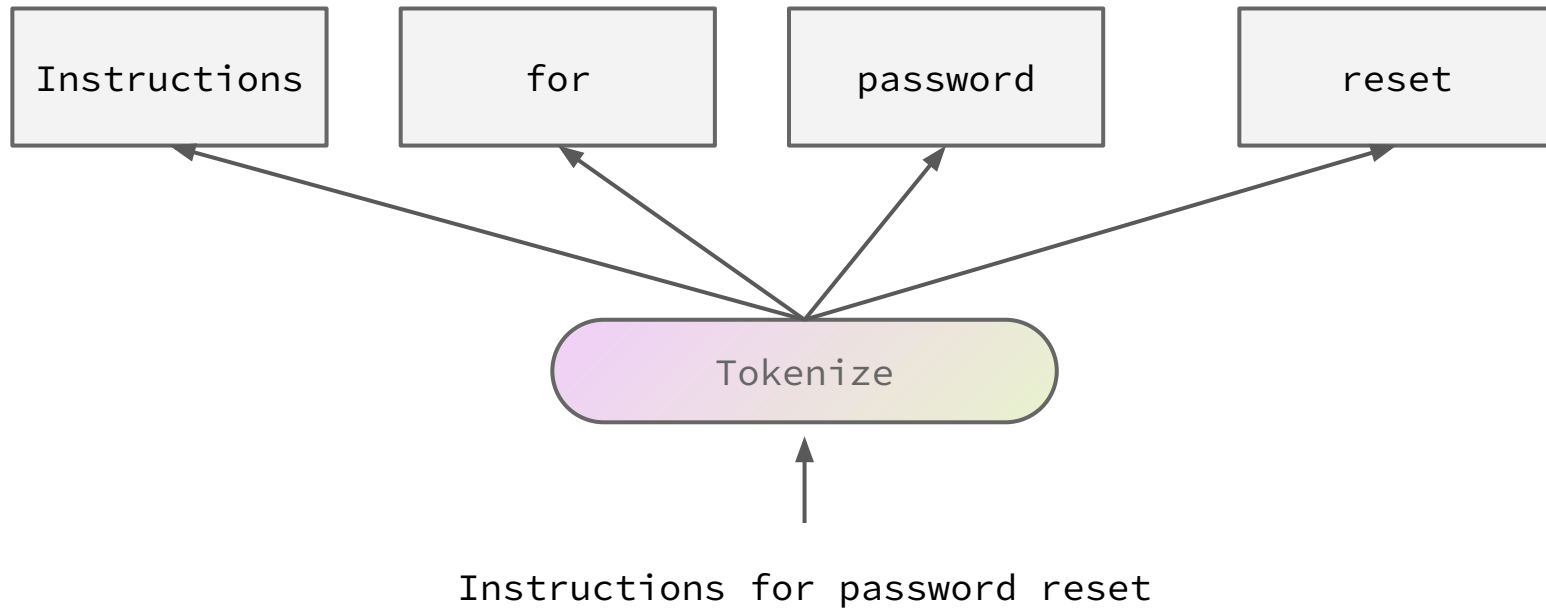
Tell me a fun fact about pandas



Though grown-up pandas are quite immense
and grand.
Their babies start out tiny, you understand!
When born they're pink, and small you see.
Just 1/900th the size their mom will be!

Tokenization

Tokenization



Tokenizer Tool

GPT-4o & GPT-4o mini

GPT-3.5 & GPT-4

GPT-3 (Legacy)

Instructions for password reset

1

Clear

Show example

3

Tokens

4

Characters

31

2

Instructions for password reset

Text in Tokenizer 1 of 2

GPT-4o & GPT-4o mini

GPT-3.5 & GPT-4

GPT-3 (Legacy)

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🍌

Sequences of characters commonly found next to each other may be grouped together: 1234567890

Clear

Show example

Tokens

53

Characters

252

Text in Tokenizer 2 of 2

Tokens

53

Characters

252

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🍌🍌🍌🍌

Sequences of characters commonly found next to each other may be grouped together:
1234567890

Text

Token IDs



Token IDs

Tokens

53

Characters

252

```
[12488, 6391, 4014, 316, 1001, 6602, 11, 889, 1236, 4128, 25, 3862, 181386, 364,
61064, 9862, 1299, 166700, 1340, 413, 12648, 1511, 1991, 20290, 15683, 290, 27899,
11643, 25, 93643, 248, 52622, 122, 279, 168191, 328, 9862, 22378, 2491, 2613, 316,
2454, 1273, 1340, 413, 73263, 4717, 25, 220, 7633, 19354, 29338, 15]
```

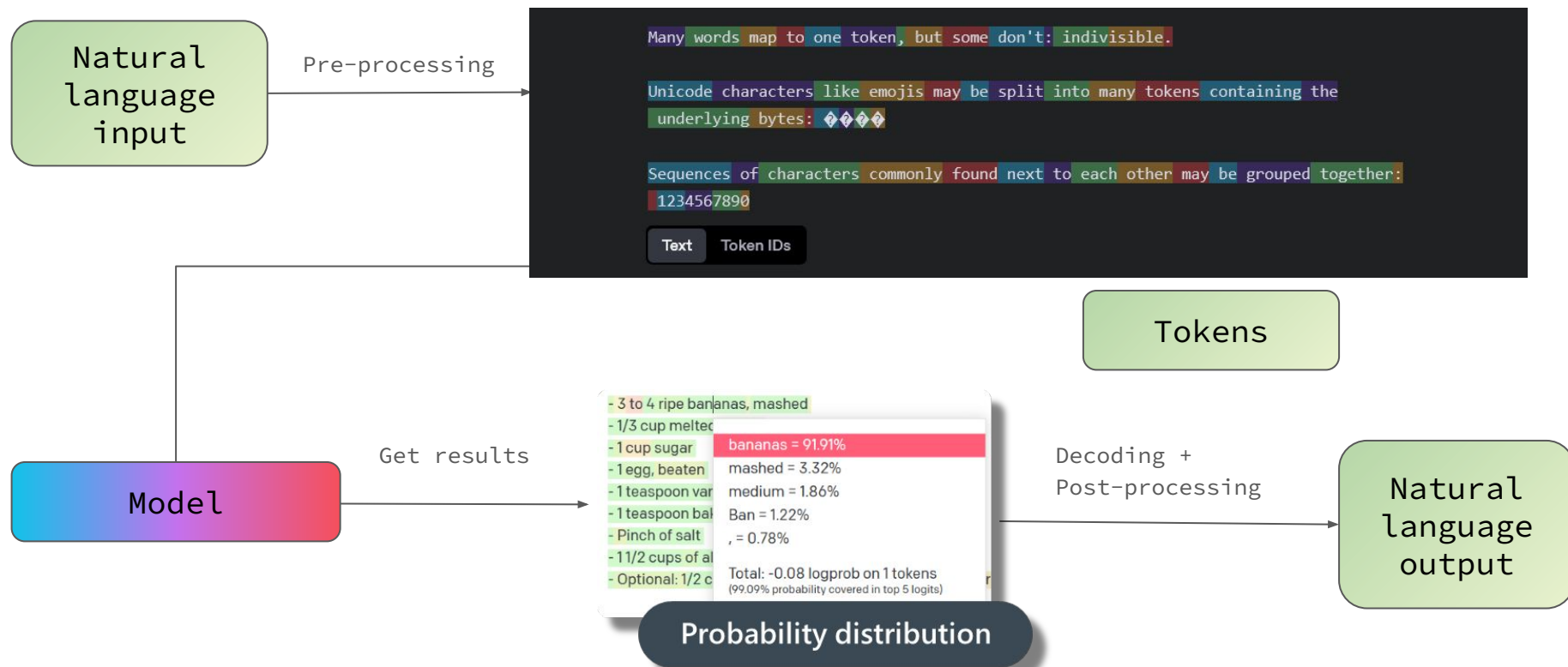
Text

Token IDs

System Message

Provide the model additional context to generate a response


How language models work



Where do we go from here?

- **Accept** that you are not going to learn everything about prompt engineering (pe) in 24 hours.
- **Inspire** others to create and discuss how to grow and build other prompts together.

Attribution

- IBM. “What Is Prompt Engineering?” IBM, 19 Dec. 2024, www.ibm.com/think/topics/prompt-engineering
- Griffiths, David. (n.d.). *Prompting Bootcamp*. O'Reilly Media - Technology and Business Training. <https://www.oreilly.com/live-events/prompting-bootcamp/0636920097098/>
-  DAIR-AI/prompt-engineering-Guide: guides, papers, lecture, notebooks and resources for prompt engineering. (n.d.). GitHub. <https://github.com/dair-ai/Prompt-Engineering-Guide>
- IBM Technology: Mixture of Experts. (2025, 14). *Manus, Vibe coding, scaling and Perplexity's AI Phone*. YouTube. <https://www.youtube.com/watch?v=Ddh3p185KhA>
- **Slide design**
 - Dair.ai. (n.d.). *Prompting Bootcamp*. O'Reilly Media - Technology and Business Training. <https://www.oreilly.com/live-events/prompting-bootcamp/0636920097098/>

Thank you



Leonardo Alvarez
Consultant | Learning
Design and Innovation...

